A Highly Accurate Machine Learning Approach for Developing Wireless Sensor Network Middleware

Remah Alshinina and Khaled Elleithy

Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, CT 06604, USA

INTRODUCTION

Despite the popularity of wireless sensor networks (WSNs) in a wide range of applications, security problems associated with them have not been completely resolved. Middleware is generally introduced as an intermediate layer between WSNs and the end user to resolve some limitations, but most of the existing middleware is unable to protect data from malicious and unknown attacks during transmission. We introduced an intelligent middleware based on an unsupervised learning technique called Generative Adversarial Networks (GANs) algorithm. GANs contain two networks: a generator (G) network and a discriminator (D) network. The G creates fake data similar to the real samples and combines it with real data from the sensors to confuse the attacker. Results illustrate that the proposed algorithm not only improves the accuracy of the data but also enhances its security by protecting data from adversaries. Data transmission from the WSN to the end user then becomes much more secure and accurate compared to conventional techniques.

PROPOSED ALGORITHM

We introduce a secure WSNs' middleware based on GANs. The proposed algorithm is capable of filtering and passing only the real data. To the best of our knowledge, it is the first time that the GANs algorithm has been used for solving the security problem in WSNs' middleware. Additionally, in the proposed contribution, WSNs' middleware applies a GAN [1] that is capable of filtering and detecting anomalies in the data.

- We propose a unique WSN middleware which can control and monitor sensor data by using intelligent, unsupervised machine learning to secure the data. The power consumption and overhead can be increased by updating and filtering unnecessary information from the sensors. This problem is addressed through the proposed unsupervised learning for middleware.
- From the given samples, the G creates fake data very similar to the real data. This fake data is combined with the real data from sensors so that the attackers cannot differentiate between them. In this case, there is no need to generate fake packets or data to confuse the attackers, which significantly decreases power consumption.

EXPERIMENTAL SETUP

GENERATOR NETWORK

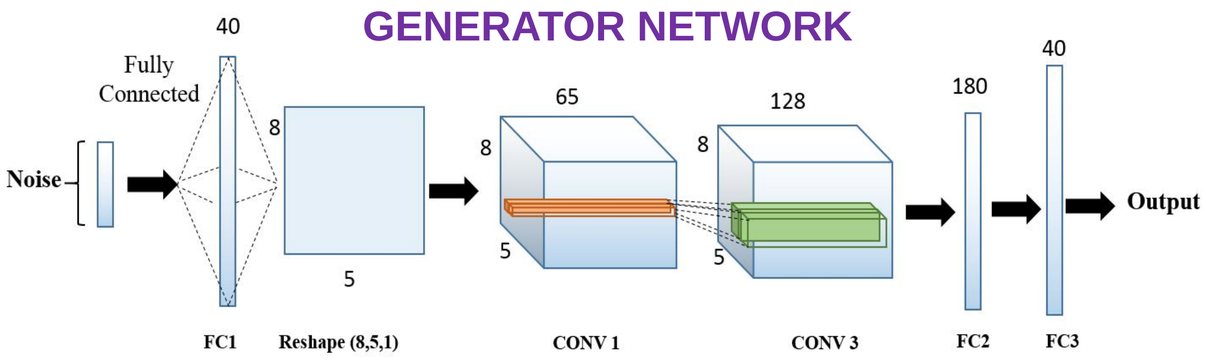

Figure 1. Experimental setup for Generator Network

DISCRIMINATOR NETWORK

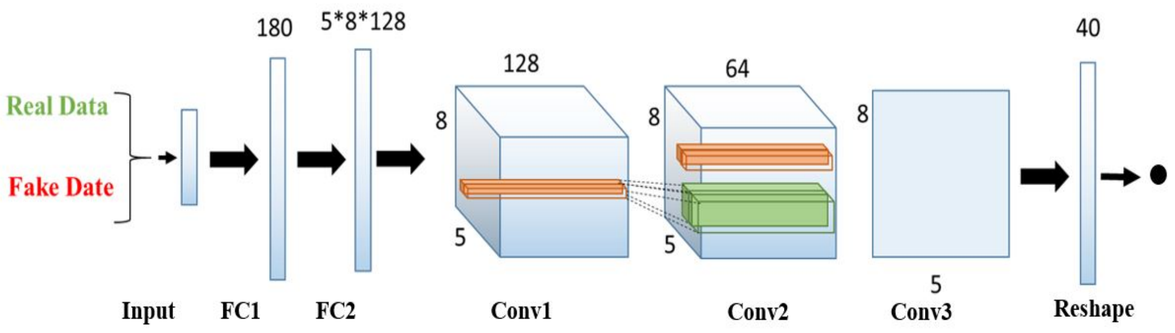

Figure 2. Experimental setup for Discriminator Network

RESULTS

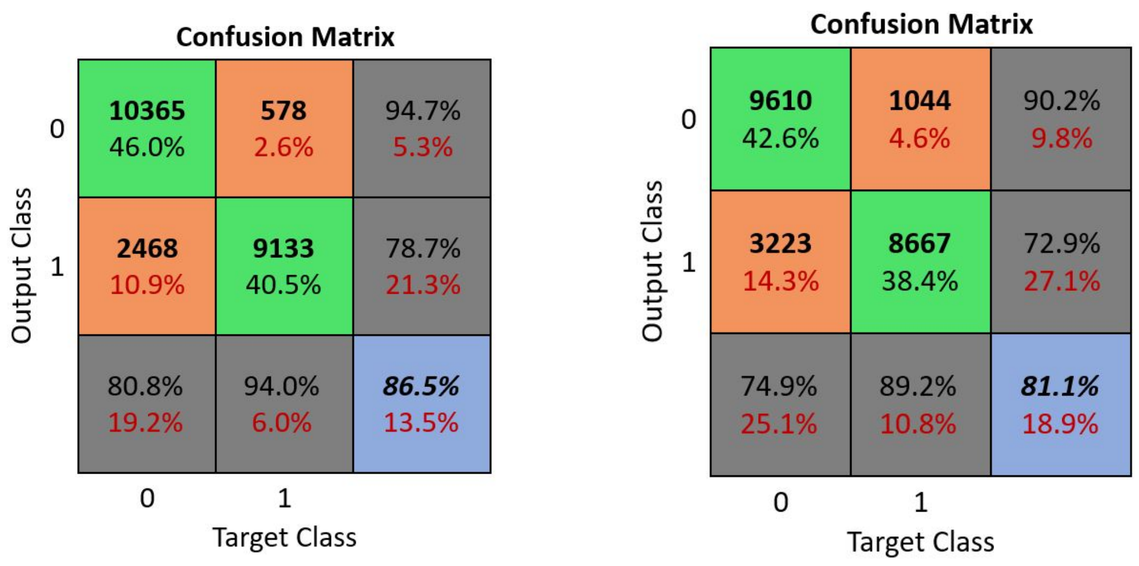

Figure 4. Left: The proposed Generator Network to generate Fake data. Right: Original Dataset (NSL-KDD).

TABLE 1. COMPARISON OF PROPOSED METHOD WITH DIFFERENT METHODS

Method	FP	FPR	F-Measure
Original Data [2]	14.3%	0.27	81.8%
Artificial Neural Network (ANN) [3]	17.4%	0.31	81.6%
Proposed Approach	**10.9%**	**0.21**	**87.2%**

TABLE 2. THE COMPARISONS OF ACCURACY RATE FOR OUR FRAMEWORK WITH DIFFERENT MACHINE LEARNING METHODS

Method	Accuracy
ANN based IDS [3]	81.2%
SVM [4]	69.52%
Decision Tree [4]	81.5%
DMNB with RP [5]	81.47%
SOM [6]	75.49%
Proposed Approach	**86.5%**

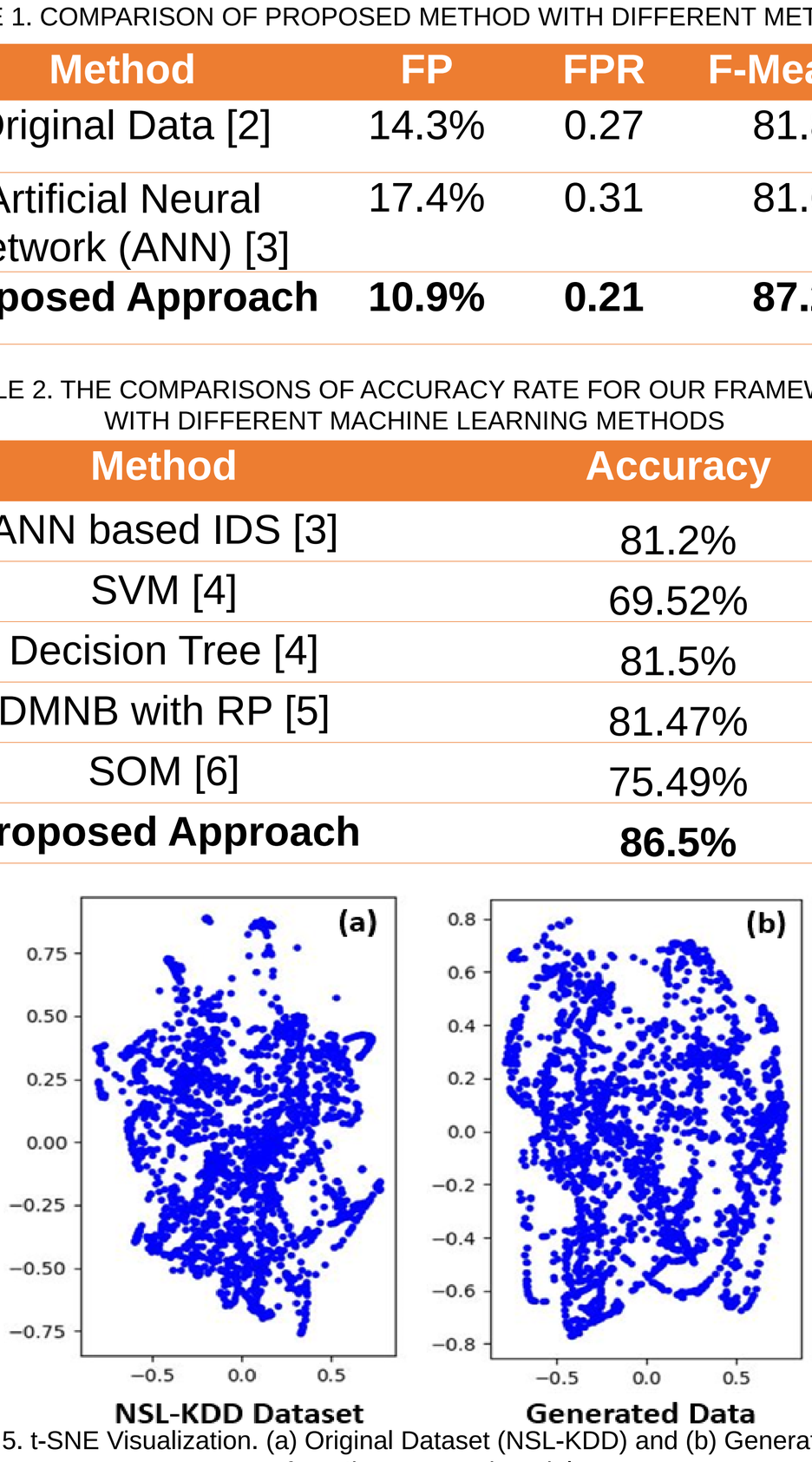

Figure 5. t-SNE Visualization. (a) Original Dataset (NSL-KDD) and (b) Generator samples from the proposed model.

CONCLUSIONS

Wireless sensor networks (WSNs) are an essential medium for the transmission of data for numerous applications. In order to address power consumption, communication, and security challenges, middleware bridges the gap between applications and WSNs. Most existing middleware does not completely address the issues that significantly impact WSNs' performance. Thus, proposes unsupervised learning for the development of WSNs middleware to provide end-to-end secure system.

- The proposed algorithm consists of a generator and a discriminator networks.
- The generator Network is capable of creating fake data to confuse the attacker and resolving imbalanced data by generating more data to balance the proportion of classes, the normal and attack data.
- We render the discriminator network to be a powerful network that can easily distinguish between two datasets, even if the fake data is very close to real samples.
- Extensive testing on the NSL-KDD dataset with different supervised learning techniques and comparisons with our generator network data shows that our generator model provides a better accuracy of 86.5% with a lower FPR.
- Results show that the proposed generator performs very well with data visualization while the original, conventional dataset NSL-KDD performed worse in t-SNE algorithm.

REFERENCES

1. I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672-2680.
2. http://www.unb.ca/cic/datasets/nsl.html. University of New Brunswick.
3. B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in *International Conference on Signal Processing and Communication Engineering Systems*, Guntur, India, 2015, pp. 92-96: IEEE.
4. Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, Canada, 2009, pp. 1-6
5. M. Panda, A. Abraham, and M. R. Patra, "Discriminative multinomial Naïve Bayes for network intrusion detection," in Sixth International Conference on Information Assurance and Security, Atlanta, GA, USA, 2010, pp. 5-10: IEEE.
6. L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod, "A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self organization map (SOM) artificial neural network," *Journal of Engineering Science and Technology,* vol. 8, no. 1, pp. 107-119, 2013.